PLOS ONE

RESEARCH ARTICLE

Prognostic function to estimate the probability of meaningful clinical improvement after surgery - Results of a prospective multicenter observational cohort study on patients with lumbar spinal stenosis

Ulrike Held[1,2☯]*, **Jakob M. Burgstaller**[1☯], **Maria M. Wertli**[1,3], **Giuseppe Pichierri**[1], **Sebastian Winklhofer**[4], **Florian Brunner**[5], **François Porchet**[6], **Mazda Farshad**[7], **Johann Steurer**[1]

1 Horten Centre for Patient Oriented Research and Knowledge Transfer, University of Zurich, Zurich, Switzerland, **2** Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland, **3** Division of General Internal Medicine, Bern University Hospital, Bern University, Bern, Switzerland, **4** Department of Neuroradiology, University Hospital Zurich, Zurich, Switzerland, **5** Department of Physical Medicine and Rheumatology, Balgrist University Hospital, Zurich, Switzerland, **6** Department of Orthopedics and Neurosurgery, Spine Center, Schulthess Clinic, Zurich, Switzerland, **7** Spine Division, Balgrist University Hospital, Zurich, Switzerland

☯ These authors contributed equally to this work.
* Ulrike.Held@uzh.ch

Check for updates

OPEN ACCESS

Citation: Held U, Burgstaller JM, Wertli MM, Pichierri G, Winklhofer S, Brunner F, et al. (2018) Prognostic function to estimate the probability of meaningful clinical improvement after surgery - Results of a prospective multicenter observational cohort study on patients with lumbar spinal stenosis. PLoS ONE 13(11): e0207126. https://doi.org/10.1371/journal.pone.0207126

Editor: Dean Chou, University of California San Francisco, UNITED STATES

Received: April 5, 2018

Accepted: October 25, 2018

Published: November 8, 2018

Copyright: © 2018 Held et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying this study is from the Lumbar Stenosis Outcome Study (LSOS) and contain sensitive patient information. Interested researchers may obtain the data through formal application to the Horten Centre Foundation (info@evimed.ch) upon publication.

Funding: We thank the Helmut Horten Foundation, the Baugarten Foundation, the Pfizer-Foundation

source: https://doi.org/10.7892/boris.123175 | downloaded: 5.5.2023

Abstract

Background

Approximately two thirds of patients with lumbar spinal stenosis (LSS) who undergo surgical treatment benefit from the surgery. The objective of this study was to derive a prognostic probability function (PPF) to identify patients with a high probability of post-surgical improvement because there is currently no method available.

Methods

In this multicenter, prospective, observational study, we collected data from eight medical centers in Switzerland in which patients underwent surgery for LSS. The endpoints were meaningful clinically important differences (MCID) in pain and disability one year after baseline. We developed a PPF named PROCESS (PostopeRative OutComE Spinal Stenosis), based on a large set of prognostic indicators extracted from the literature. The PPF was derived using data from a random subset of two thirds of the patients and validated in the remaining third. We addressed overfitting by shrinking the regression coefficients. The area under the ROC curve (AUC) and calibration determined the accuracy of the PPF.

Results

In this study, 452 LSS patients received surgery. 73% of the 300 patients in the derivation subset reached an MCID in pain and 68% reached an MCID in disability. The corresponding

for geriatrics and research in geriatrics, the Symphasis Charitable Foundation, and the OPO Foundation for their support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

values were 70% and 63% in the validation subset, respectively. In the derivation subsample, the AUC was 0.64 (95% CI 0.57 to 0.71) for of the PPF predicting MCID in pain and 0.71 (0.64 to 0.77) for MCID in disability, after shrinkage. The corresponding numbers were 0.62 (0.52 to 0.72) and 0.70 (0.60 to 0.79) in the validation subsample, and the PPF showed good calibration.

Conclusions

Surgical treatment for patients with lumbar spinal stenosis is being performed with increasing frequency. PROCESS is conditional on the individual pattern of preoperatively available prognostic indicators, and may be helpful for clinicians in counselling patients and in guiding the discussion on individual treatment decision in the era of personalized medicine.

Introduction

Decompression surgery is a treatment option for patients with lumbar spinal stenosis. Surgical treatment is recommended for patients with moderate or severe clinical manifestations and no meaningful improvement following conservative treatment, such as physiotherapy and/or epidural steroid injections. Surveys among surgeons using standardized clinical cases revealed a lack of consensus among clinical experts on the indications for surgery [1], and the wide variation in surgical rates among hospital referral regions in the USA may be explained by this lack of consensus [2].

There is broad agreement that decompressive surgery offers an advantage for about two thirds of patients within the first four years after surgery compared to non-surgical treatment [3–6], while one third of patients report no meaningful improvement after surgery [4, 5]. This difference in outcome declines between surgical and non-surgical treatment five to ten years after surgery [7–9]. A tool for identifying patients with a high probability of post-surgical improvement would be valuable for patients and physicians in decision-making regarding surgery, and could reduce the rate of unnecessary operations.

The aim of this study was to develop a PPF for estimating the likelihood of post-surgical improvement in patients with lumbar spinal stenosis at one year follow-up, conditional on a set of prognostic indicators measured at baseline.

Materials and methods

The Lumbar Stenosis Outcome Study (LSOS) is a prospective cohort study investigating the effectiveness of various treatment options in patients with symptomatic lumbar spinal stenosis [10]. Participation in the study had no influence on the treatment of the patients, all treatment decisions were left to the patient and physician. 841 patients were recruited from December 2010 to December 2015, and were followed-up for three years.

The present study is reported according to STROBE (Statement for reporting cohort studies) guidelines [11] and TRIPOD (Statement for studies reporting clinical prediction models) guidelines [12] (S1 and S2 Files).

Inclusion and exclusion criteria

The following characteristics were required for study eligibility: age $\geq$50 years, uni- or bilateral neurogenic claudication, verified stenosis of the lumbar spinal canal determined by magnetic

resonance imaging (MRI), life expectancy ≥1 year, ability to give informed consent, availability for follow-up, and ability to complete questionnaires in the German language. We excluded patients with cauda equina syndrome requiring urgent surgery, current fracture, infection, significant deformity (>15˚ lumbar scoliosis) of the lumbar spine, current enrollment in another spine-related treatment study, and clinically relevant peripheral arterial disease (as confirmed by a vascular specialist in patients without palpable lower limb pulse).

Eligibility criteria for inclusion in the analysis

All patients who underwent surgery within six months of enrollment and with 12 months of follow-up data were included in the analysis.

Surgical procedures and radiological classification

The surgery consisted of a standard open posterior lumbar laminotomy with or without instrumentation of the affected level(s). The surgeon's discretion determined the decision to proceed with a laminotomy using unilateral technique, to decompress the contralateral recess, or to take a midline approach with bilateral laminotomy. Fusion surgery included implantation of pedicle screws with rods, plus intersomatic fusion and cage(s) at the affected level(s), in addition to decompression surgery. Additional fusion, single or multi-level decompression, or the use of an operating microscope was based on the surgeon's discretion. The procedures were done or supervised by senior neuro- or orthopedic surgeons with more than ten years of experience after board-certification, and each patient's MRI was independently evaluated by two senior radiologists.

Data collection and follow-up

Baseline data was taken from interviews and recorded by a study coordinator. All other questionnaires were self-administered by the patients themselves. Data were collected at baseline and after 12 months.

Outcome measures

Spinal Stenosis Measure (SSM): The SSM is an instrument specifically developed and validated for spinal stenosis patients by Stucki and colleagues [13], and both measures symptom severity and quantifies disability in patients with lumbar spinal stenosis. The SSM is recommended by the North American Spine Society (NASS) and is used in many different studies of lumbar spinal stenosis [14–17]. Scores range from 1–5 and 1–4 (best-worst) for SSM symptoms score and SSM function score. A minimal clinically important difference (MCID) in SSM symptoms score is defined by an improvement of 0.48 points from baseline to 12 months, and an improvement of 0.52 defines an MCID in SSM function score [18].

Development of a PPF and choice of prognostic indicators

A PPF derives the probability of a future event based on a set of prognostic indicators defined at a specified point in time [19]. In this application, the future event was MCID one year after baseline in SSM symptoms score or SSM function score. We studied the literature to generate a list of prognostic indicators [20–23]. The dichotomous and continuous parameters measured at baseline and available in the LSOS database were included in the PROCESS (PostopeRative OutComE Spinal Stenosis) PPF. The list of prognostic indicators is summarized in Table 1.

Table 1. List of 21 prognostic indicators used in the prognostic probability function for meaningful clinically important difference (MCID) in SSM symptoms score and SSM function score one year after baseline.

Prognostic indicator	
Dichotomous	*Unfavorable*
Age	≥75 years [21]
Gender	female
BMI	≥30 kg/m^2 [24]
Current smoker	yes
Civil status	living alone, or single/divorced/widowed and living in a nursing/ residential home
Formal education	compulsory school only
Coxarthrosis or gonarthrosis	yes
Coronary heart disease or heart insufficiency	yes
Asthma or COPD	yes
Parkinson's disease or peripheral neuropathy	yes
Walking ability	being able to walk only up to 200 m
Low back pain	yes
Duration of symptoms	≥6 months [25]
Preoperative analgesic use within 3 months before baseline	yes
Previous lumbar surgery	yes
Number of decompressed levels	>1 level
Radiological parameters	Antero-posterior diameter of dural sac (APD) >6 mm or cross sectional area >70 mm^2 [22]
Depression (on HADS depression scale)	≥8 points [26]
Continuous	*Range*
Quality of life (EQ5D-3L scale)	0 (worst)– 100 (best)
Baseline SSM symptoms score	1 (best)– 5 (worst)
Baseline SSM function score	1 (best)– 4 (worst)

BMI = body mass index; COPD = chronic obstructive pulmonary disease; HADS = Hospital Anxiety and Depression Scale; SSM = Spinal Stenosis Measure

https://doi.org/10.1371/journal.pone.0207126.t001

Patient scenarios

The PPF's usability in terms of pain and disability was also assessed by characterizing two scenarios, one with a very favorable constellation of prognostic indicators, and the second with a disadvantageous constellation of prognostic indicators. We calculated the estimated probabilities of MCID in these two scenarios.

Scenario 1 (favorable). Male patient, all favorable prognostic indicators present (Table 1), age <75 years, BMI <30 kg/m^2, non-smoking, not living alone or single/divorced/widowed and living in a nursing/residential home, no low education, no cox- or gonarthrosis, coronary heart disease or heart insufficiency, no asthma or chronic obstructive pulmonary disease (COPD), no Parkinson's disease or peripheral neuropathy, able to walk more than 200 m, no low back pain, duration of symptoms <6 months, analgesic use within 3 months before baseline, no previous lumbar surgery, surgery on one single lumbar spinal level, narrowing of the cross sectional area (≤70 mm^2) or the diameter of the dural sac (≤6 mm), no depression, average quality of life score of 50 points, baseline SSM symptoms and function scores both 3.5.

Scenario 2 (unfavorable). Female patient, all favorable prognostic indicators absent (Table 1), average quality of life score of 50, baseline SSM symptoms and function scores both 2.

Statistical methods

Thorough development and validation of a PPF is important [27]. For that reason, patients were randomly split once into a *derivation subsample* (2/3 of the patients) for development of the PPF and a *validation subsample* (1/3 of the patients) for validation of the function, in order to determine the validity of results for new patients [28]. Descriptive statistics included median and interquartile ranges for continuous variables, and counts and percentages of total for categorical variables. Corresponding Wilcoxon and chi-squared tests were used to compare the two subsamples.

There was missing data for some of the patients for some of the prognostic indicators. These were filled using 10-fold multiple imputation based on chained equations [29], retaining the information about the derivation and validation subsamples.

The two binary outcome variables, MCID in SSM symptoms and SSM function were addressed with logistic regression models fitted to each outcome in each of the ten imputed derivation subsamples including all 21 prognostic indicators.

PPF models with a large number of prognostic indicators tend to describe optimally the data under study, but predictions for *new* subjects will perform less well. To address this phenomenon, called overfitting, the regression coefficients of the PPF can be shrinked towards zero by multiplying with a global shrinkage factor. E.g., a coefficient of 0.8 becomes 0.72 (= 0.8*0.9) if the shrinkage factor was 0.9. We derived a global shrinkage factor for the estimated regression coefficients using the *dfbeta*-method [30]. A global shrinkage factor for the model addressing MCID in SSM symptoms and a global shrinkage factor for MCID in SSM function was calculated in each of the ten imputed derivation subsamples and then averaged, resulting in one global shrinkage factor for each outcome.

Pooled regression coefficients were calculated from the ten derivation subsamples following Rubin's rule [31], and the two global shrinkage factors were applied. Original and shrunken regression coefficients are summarized as odds ratios with 95% confidence intervals (CI). The pooled and shrunken regression coefficients were applied to each of the ten multiply imputed validation subsamples, resulting in predictions of the probability for MCID in SSM symptoms and SSM function scores.

After the derivation of a PPF, one would like to know how well the predicted probability for MCID (continuous between 0 and 1) corresponds to the actual observed MCID-status (0 or 1) in SSM symptoms and SSM function. This can be measured with the discriminative ability of the PPF, as well as with its calibration. A **r**eceiver **o**perating **c**haracteristic (ROC) plot displays the true positive rate against the false positive rate for consecutive cut-offs for the predicted probability. The **a**rea **u**nder this ROC **c**urve (AUC) with 95% CI is calculated to assess the discriminative ability of the PPF. Calibration is another important property of a probability function, and it measures the agreement between observed outcomes and predictions. We used calibration plots, for 10-fold imputed derivation and validation subsamples.

All analyses were conducted using R for Windows [32], using the packages *dplyr*, *MASS*, *mice*, *mitools*, *openxlsx*, *PresenceAbsence*, *pROC*, *rms*, *rpart*, *shrink*, and *tableone*. Our work was conducted following the concept of reproducible research and the R-code is available upon request [33].

Sample size

The sample size was calculated for the development of a PPF for patients undergoing spine surgery, one year after surgery [10]. For sample size calculation, we anticipated that 60% of the included 841 patients (= 505 patients) with verified diagnosis would undergo surgery. Actually, 543 patients underwent surgery during the follow-up, 498 of these within the first six months after baseline and 452 of these had a follow-up of at least 12 months.

We anticipated that two thirds of the patients would show a clinically relevant improvement one year after surgery. Based these assumptions, the number of prognostic indicators in the logistic regression model in the derivation set may be up to 20 following the rule of 10 outcome events per predictor variable (EPV). According to Vittinghoff and McCulloch [34], this rule can be relaxed to 5 to 9 EPVs allowing for up to 22 prognostic indicators if 9 EPVs are taken as threshold.

Ethical approval

This multi-centre cohort study was conducted in compliance with all international laws and regulations as well as any applicable guidelines. Written informed consent to participate in the study has been obtained from participants. The study was approved by the independent Ethics Committee of the Canton Zurich (KEK-ZH-NR: 2010-0395/0).

Results

Patient characteristics

Four hundred and fifty-two patients who received surgery for lumbar spinal stenosis within six months of baseline had 12-month follow-up data (Fig 1). 300 were randomly selected for

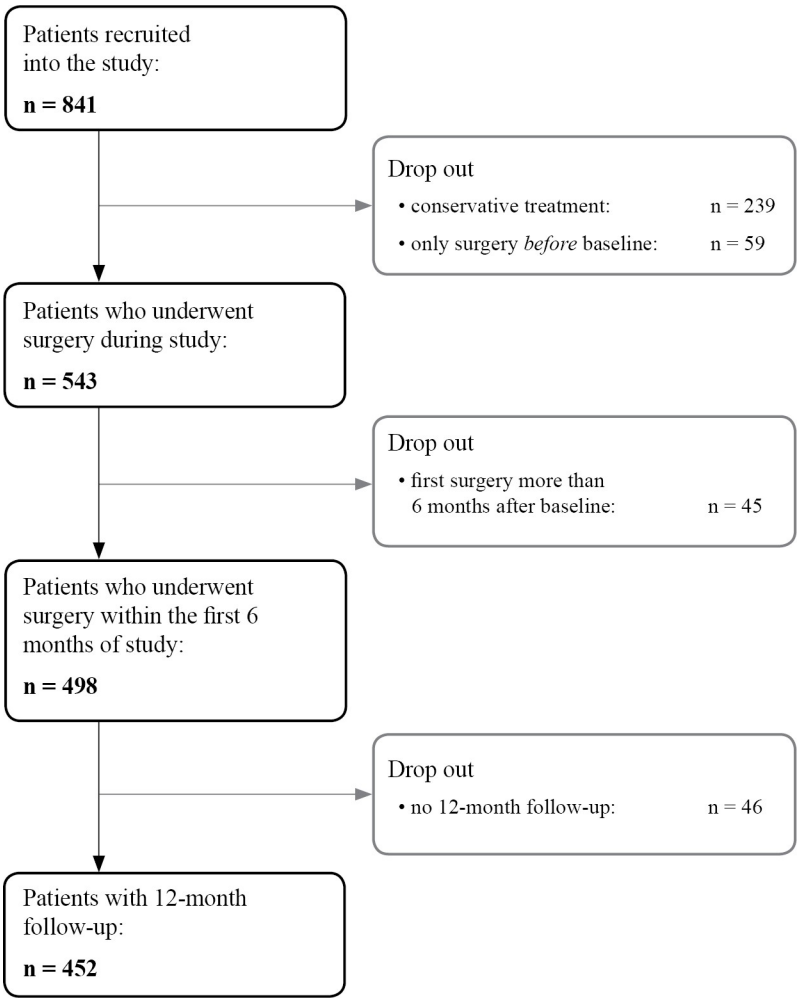

Fig 1. Patient flow chart.

https://doi.org/10.1371/journal.pone.0207126.g001

the derivation subsample, and the remaining 152 patients were allocated to the validation subsample. 141 patients in the derivation subsample (47%) were 75 years or older, while 73 (48%) of patients in the validation subsample were in the same age range. 51% and 52%, respectively, were female. Details for all prognostic indicators are shown in Table 2. There were no differences between the subsamples across all variables.

MCID in SSM symptoms and SSM function scores

73% of the patients in the derivation subsample and 70% of patients in the validation subsample reached an MCID in SSM symptoms score after surgery. The percentages for MCID in SSM function score were 68% and 63%, respectively.

Intra- and postoperative complications

Eighteen (6.0%) patients of the derivation subsample suffered an injury of the dura during surgery, the corresponding numbers were 9 (5.9%) in the validation subsample. No patient in the derivation subsample and three patients (2.0%) in the validation subsample experienced epidural venous bleeding.

In the derivation subsample, 5 (1.7%) patients had a wound infection, other complications (e.g., urosepsis, hemorrhage, wound healing deficit) were experienced by 25 (8.3%) patients. In the validation subsample, no wound infections were observed, and 15 (9.9%) patients had other complications. Twenty-two (7.3%) patients had a reoperation in the derivation subsample and 15 (9.9%) patients in the validation subsample.

Table 2. Preoperative baseline characteristics.

	Derivation data set	Validation data set	p-value
	N = 300	N = 152	
Age ≥75 years, No. (%)	141 (47.0)	73 (48.0)	0.915
Female gender, No. (%)	152 (50.7)	79 (52.0)	0.870
BMI ≥30 kg/m^2, No. (%)	83 (27.7)	47 (30.9)	0.540
Current smoker, No. (%)	50 (16.7)	24 (15.9)	0.929
Living alone, or single/divorced/widowed and living in nursing home, No. (%)	95 (31.7)	56 (36.8)	0.319
Compulsory school only, No. (%)	70 (23.5)	45 (29.6)	0.196
Coxarthrosis or gonarthrosis, No. (%)	105 (38.0)	54 (36.7)	0.873
Coronary heart disease or heart insufficiency, No. (%)	19 (6.9)	10 (6.8)	1
Asthma or COPD, No. (%)	28 (10.0)	19 (12.8)	0.465
Parkinson's disease or peripheral neuropathy, No. (%)	7 (2.5)	6 (4.1)	0.552
Being able to walk only up to 200m, No. (%)	209 (69.9)	99 (65.1)	0.357
Low back pain, No. (%)	263 (88.0)	133 (88.1)	1
Duration of symptoms ≥6 months, No. (%)	182 (61.1)	87 (58.0)	0.600
Preoperative analgesic use within 3 months before baseline, No. (%)	242 (80.9)	117 (78.5)	0.633
Previous lumbar surgery, No. (%)	27 (9.0)	22 (14.5)	0.108
More than one decompressed level, No. (%)	181 (60.5)	82 (55.8)	0.392
Diameter of the dural sac (APD) >6 mm or cross sectional area >70 mm^2 (%), No. (%)	52 (17.3)	34 (22.4)	0.245
Depression on HADS scale ≥8, No. (%)	56 (18.7)	26 (17.2)	0.793
Quality of life on EQ5D actual health status, median [IQR]	64.0 [40.0, 80.0]	63.5 [40.0, 80.0]	0.818
Baseline SSM symptoms score, median [IQR]	3.1 [2.7, 3.6]	3.1 [2.9, 3.5]	0.912
Baseline SSM function score, median [IQR]	2.2 [1.8, 2.8]	2.2 [1.8, 2.8]	0.808

BMI = body mass index; COPD = chronic obstructive pulmonary disease; HADS = Hospital Anxiety and Depression Scale; SSM = Spinal Stenosis Measure

https://doi.org/10.1371/journal.pone.0207126.t002

Multiple imputation and shrinkage factors

There were missing values for 18 of the prognostic indicators. The percentage of missing values varied between 0.2% in walking ability, and 6.4% for the coxarthrosis/gonarthrosis variable. Ten-fold multiple imputation was applied.

The continuous prognostic indicators baseline SSM symptoms score and SSM function score, and EQ-5D actual health status were entered in a linear as well as in a quadratic fashion and the residual plot was in favor of the linear effect for all three of them.

The resulting average global shrinkage factors were 0.47 for MCID in SSM symptoms score and 0.60 for MCID in SSM function score.

Pooled regression coefficients expressed as log odds ratios and 95% CIs are summarized in Table 3 (MCID in SSM symptoms score as outcome) and Table 4 (MCID in SSM function score as outcome). The shrinked regression coefficients are also displayed for each outcome. These were the final coefficients, and were used for calculating the probability of MCID in SSM symptoms score and SSM function score following surgery.

Discriminative ability of PROCESS, calibration and validation

In the derivation subsample, the discriminative ability as measured with the AUC of the PPF was 0.64 (95% CI 0.57 to 0.71) for the MCID in SSM symptoms score after shrinkage. The AUC was 0.62 (0.52 to 0.72) in the validation subsample. The corresponding values were 0.71

Table 3. PROCESS: Estimated coefficients and their 95% confidence intervals, p-values, and shrinked coefficients for the SSM symptoms score. The shrinkage factor was 0.47.

MCID in SSM symptoms	Coefficients = log odds ratios	Lower bound of 95% CI	Upper bound of 95% CI	p-value	Shrinked coefficients
(Intercept)	-1.212	-3.878	1.455	0.372	-0.565
Age	-0.460	-1.090	0.171	0.152	-0.214
Gender	0.108	-0.568	0.785	0.753	0.050
Body mass index	-0.443	-1.118	0.232	0.197	-0.207
Current smoker	-0.021	-0.853	0.811	0.960	-0.010
Civil status	-0.178	-0.874	0.519	0.616	-0.083
Formal education	-0.232	-0.924	0.460	0.510	-0.108
Coxarthrosis or gonarthrosis	-0.399	-1.041	0.243	0.221	-0.186
Coronary heart disease or heart insufficiency	-1.221	-2.348	-0.093	0.034	-0.569
Asthma or COPD	0.612	-0.609	1.833	0.324	0.285
Parkinson's disease or peripheral neuropathy	-1.218	-3.115	0.679	0.206	-0.568
Walking ability	-0.189	-0.939	0.561	0.621	-0.088
Low back pain	-0.439	-1.405	0.528	0.372	-0.205
Duration of symptoms	0.004	-0.618	0.626	0.990	0.002
Preoperative analgesic use within 3 months before baseline	-0.070	-0.864	0.724	0.862	-0.033
Previous lumbar surgery	0.050	-0.958	1.057	0.923	0.023
Number of decompressed levels	-0.269	-0.882	0.344	0.389	-0.125
Antero-posterior diameter of dural sac (APD) >6 mm or cross sectional area >70 mm^2	-0.694	-1.421	0.032	0.061	-0.324
Depression (on HADS depression scale)	-1.091	-1.872	-0.311	0.006	-0.509
Quality of life (EQ5D-3L scale)	0.001	-0.013	0.015	0.882	0
Baseline SSM symptoms score	1.396	0.761	2.032	<0.001	0.652
Baseline SSM function score	-0.174	-0.808	0.460	0.589	-0.081

CI = confidence interval; COPD = chronic obstructive pulmonary disease; HADS = Hospital Anxiety and Depression Scale; SSM = Spinal Stenosis Measure

https://doi.org/10.1371/journal.pone.0207126.t003

Table 4. PROCESS: Estimated coefficients and their 95% confidence intervals, p-values, and shrinked coefficients for the SSM function score. The shrinkage factor was 0.60.

MCID in SSM function	Coefficients = log odds ratios	Lower bound of 95% CI	Upper bound of 95% CI	p-value	Shrinked coefficients
(Intercept)	-1.054	-3.692	1.584	0.432	-0.633
Age	-0.603	-1.233	0.028	0.061	-0.362
Gender	-0.002	-0.670	0.665	0.994	-0.001
Body mass index	-0.957	-1.634	-0.281	0.006	-0.574
Current smoker	0.467	-0.370	1.304	0.273	0.280
Civil status	0.592	-0.133	1.316	0.109	0.355
Formal education	0.113	-0.607	0.834	0.757	0.068
Coxarthrosis or gonarthrosis	-0.711	-1.364	-0.058	0.033	-0.427
Coronary heart disease or heart insufficiency	-0.742	-1.834	0.351	0.182	-0.445
Asthma or COPD	-0.722	-1.769	0.326	0.176	-0.433
Parkinson's disease or peripheral neuropathy	-1.103	-2.930	0.725	0.234	-0.662
Walking ability	0.108	-0.676	0.892	0.786	0.065
Low back pain	-1.127	-2.111	-0.142	0.025	-0.676
Duration of symptoms	-0.046	-0.672	0.580	0.885	-0.028
Preoperative analgesic use within 3 months before baseline	-0.366	-1.168	0.435	0.369	-0.220
Previous lumbar surgery	-0.987	-1.951	-0.022	0.045	-0.592
Number of decompressed levels	-0.377	-1.001	0.247	0.235	-0.226
Antero-posterior diameter of dural sac (APD) >6 mm or cross sectional area >70 mm^2	-0.571	-1.322	0.180	0.136	-0.343
Depression (on HADS depression scale)	-0.552	-1.350	0.247	0.175	-0.331
Quality of life (EQ5D-3L scale)	0.010	-0.004	0.025	0.169	0.006
Baseline SSM symptoms score	-0.239	-0.834	0.357	0.431	-0.143
Baseline SSM function score	2.038	1.312	2.764	<0.001	1.223

CI = confidence interval; COPD = chronic obstructive pulmonary disease; HADS = Hospital Anxiety and Depression Scale; SSM = Spinal Stenosis Measure

https://doi.org/10.1371/journal.pone.0207126.t004

(0.64 to 0.77) in the derivation subsample and 0.70 (0.60 to 0.79) in the validation subsample for the MCID in SSM function score.

The corresponding ROC curves are displayed in Fig 2 by applying the PPF with shrinked coefficients to the ten derivation (black lines) and the ten validation (grey lines) subsamples resulting from the multiple imputation. The left panel shows the MCID in SSM symptoms score, and the right panel shows MCID in SSM function score.

The calibration plots in Fig 3 show that there was good overall calibration of the PPFs when probabilities for MCID were displayed against bins of observed probabilities for MCID. The upper panels show results for SSM symptoms score, and the lower panels show results for SSM function score.

The information whether fusion was added or not to decompression surgery was entered as an additional variable to the PPF. The resulting AUC of the PPF in the derivation subsample was 0.64 (95% CI 0.57 to 0.71) for the MCID in SSM symptoms score after shrinkage and 0.62 (0.52 to 0.72) in the validation subsample. The corresponding values for the MCID in SSM function score were 0.71 (0.64 to 0.77) and 0.68 (0.59 to 0.77).

Specific patient scenarios

The estimated probability for MCID after surgery in SSM symptoms score was 81% in the favorable scenario (Scenario 1), and 9% in the unfavorable scenario (Scenario 2). For the

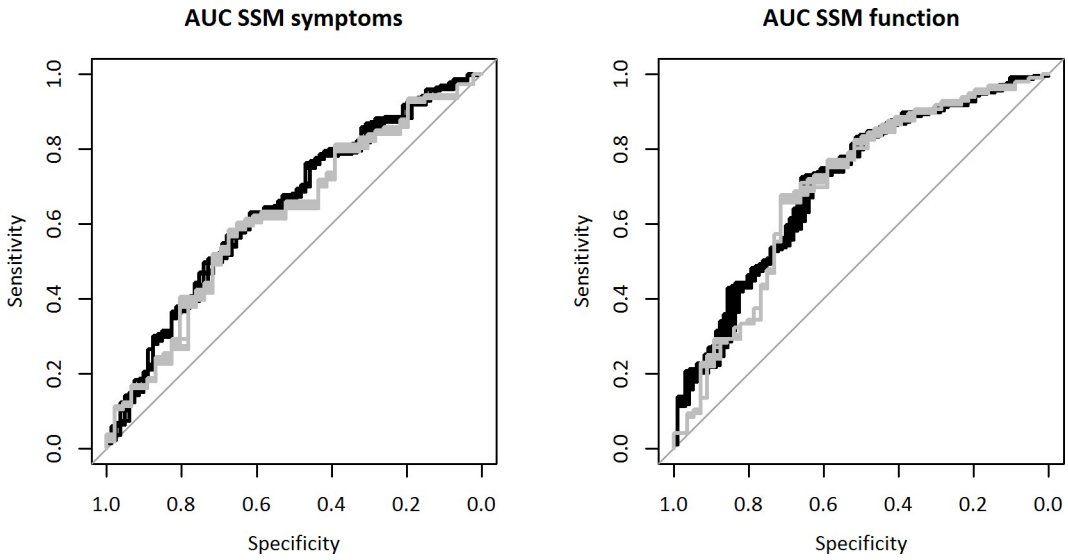

Fig 2. Receiver operating characteristic curves for MCID in SSM symptoms score and SSM function score. · Black lines show results from ten imputed derivation subsamples, grey lines show results from ten imputed validation subsamples. The discriminative ability of the PPF was 0.64 (derivation = black lines) and 0.62 (validation = grey lines) for SSM symptoms score (left panel). The corresponding values were 0.71 and 0.70 for SSM function score as outcome (right panel). MCID = meaningful clinically important difference; SSM = Spinal Stenosis Measure.

https://doi.org/10.1371/journal.pone.0207126.g002

MCID in SSM function score, the first scenario resulted in an estimated probability of success of 97%, while it was only 6% in the second scenario.

We demonstrate how these two scenarios lead to the aforementioned probabilities in S1 and S2 Tables. The probabilities of every other constellation of prognostic indicators can be calculated online at www.evimed.ch/PROCESS.

Discussion

We derived and validated PROCESS in a population of 452 patients with lumbar spinal stenosis in order to estimate the probability of reaching a minimal clinically important difference (MCID) one year after baseline. The discriminative ability and calibration of the PPF for MCID in SSM function score was better than that for SSM symptoms score. Approximately two thirds of the patients benefitted from spinal surgery, however, preoperative prognostic indicators had a large impact on individual outcomes. High baseline pain or functional impairment levels were among the strongest indicators positively associated with MCID in symptoms or function. Depression, low back pain, and previous lumbar surgery were negatively associated. Estimated probabilities of MCID varied, and ranged from 6% to 97%.

The authors of current treatment guidelines identified limited evidence to recommend surgical treatment for patients with lumbar spinal stenosis [35], and did not address the importance of prognostic indicators in the treatment decision. Several prognostic indicators associated with clinically meaningful improvement were identified in a systematic review. These included better reported walking capacity, better self-rated health, and shorter symptom duration [22]. Indicators for an unfavorable outcome after surgery were cardiovascular comorbidity, low back pain, and higher outcome expectations [22]. The majority of the original studies included in the systematic review were of "low quality" and based on small patient samples, likely leading to overly simplified prediction models based on a single or a few prognostic indicators. More recently, several studies identified that a higher degree of baseline disability is

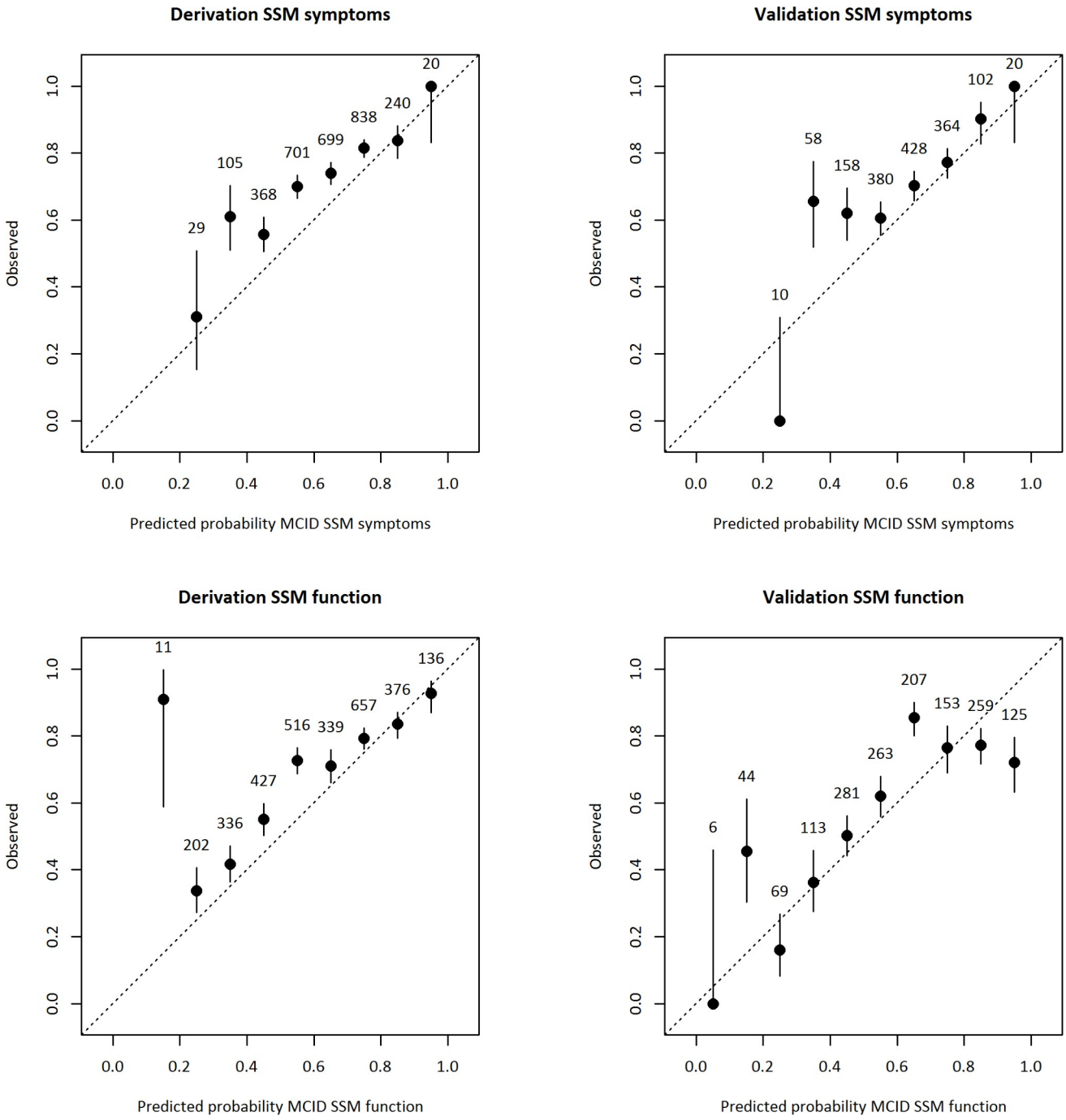

Fig 3. Calibration plots of observed versus predicted probabilities for MCID in SSM symptoms score and SSM function score. The left panel shows results from the derivation subsample, and the right panel shows results from the validation subsample. Overall, there was good calibration of the PPF when probabilities for MCID were displayed against bins of observed probabilities for MCID, shown in upper panels for SSM symptoms score and lower panels for SSM function score. MCID = meaningful clinically important difference; SSM = Spinal Stenosis Measure; MI = multiple imputation.

https://doi.org/10.1371/journal.pone.0207126.g003

associated with increased improvement of functional outcome [20, 36–38], while smoking [36, 39, 40] and psychiatric disease [20, 41] were associated with an unfavorable outcome. Athiviraham et al.[20] reported that higher BMI was associated with less functional improvement, while Pearson et al. [36] reported no difference in function between patients with BMI greater

than or equal to 30 and those with a BMI below 30. A few recently published studies considered only a limited patient sample [20, 21, 37, 38] and did not investigate the influence of the indicators on established clinically meaningful improvement [20, 37, 38].

In PROCESS, all available prognostic indicators previously identified were simultaneously included. Given the serious problem of overfitting in the development of PPFs in a large database, we deliberately refrained from an additional selection of parameters collected in the LSOS database, as this would have resulted in optimism regarding the model's predictive performance in new patients [42]. To address optimism in PROCESS, shrinkage was applied to the regression coefficients and the final models were validated in a random sample of one third of patients previously withheld from the analysis.

The discriminative ability of PROCESS was not altered by the inclusion of the information whether fusion was added or not to decompression surgery.

This study has several strengths. The data were collected prospectively in multiple study centers, and the disease-specific questionnaires SSM symptoms score and SSM function score were used to measure pain and disability. We also applied advanced methodology to obtain a robust PPF using multiple imputation techniques and shrinkage. The performance of the PPF was measured with an unused validation portion of the data set.

A weakness of our study is the fact that the prognostic indicators "coronary heart disease or cardiac insufficiency", "asthma or COPD", and "Parkinson's disease or peripheral neuropathy" happened to have a low prevalence in our data set. Another weakness is that not all risk factors published by Aalto et al. [21, 22] and Athiviraham et al. [20] were collected in the LSOS study: we had no information on preoperative scoliosis (an exclusion criterion in our study when $>15^{\circ}$), income, or outcome expectations.

Our method provided clinicians with individualized estimates of success probability with respect to pain and functional improvement that was both easy to understand and simple to communicate to the patient.

Surgical treatment for patients with lumbar spinal stenosis is being performed with increasing frequency [43], leading to higher costs for the health care system. PROCESS is conditional on the individual pattern of preoperatively available prognostic indicators, and may be helpful for clinicians in counselling patients and in guiding the discussion on individual treatment decision in the era of personalized medicine.

Supporting information

S1 Table. Use of the PROCESS prognostic probability function for the MCID in SSM symptoms outcome: Favorable and unfavorable constellation as described in the Methods and Results sections.
(DOCX)

S2 Table. Use of the PROCESS prognostic probability function for the MCID in SSM function outcome: Favorable and unfavorable constellation as described in the Methods and Results sections.
(DOCX)

S1 File. STROBE Statement—checklist of items that should be included in reports of observational studies.
(DOCX)

S2 File. TRIPOD checklist: Prediction model development.
(DOCX)

Author Contributions

Conceptualization: Ulrike Held, Jakob M. Burgstaller, Johann Steurer.

Data curation: Jakob M. Burgstaller, Giuseppe Pichierri.

Formal analysis: Ulrike Held, Jakob M. Burgstaller.

Funding acquisition: Johann Steurer.

Investigation: Ulrike Held, Sebastian Winklhofer, Florian Brunner, François Porchet, Mazda Farshad, Johann Steurer.

Methodology: Ulrike Held, Jakob M. Burgstaller, Maria M. Wertli, Johann Steurer.

Project administration: Johann Steurer.

Software: Ulrike Held, Jakob M. Burgstaller.

Validation: Ulrike Held, Jakob M. Burgstaller.

Visualization: Giuseppe Pichierri.

Writing – original draft: Ulrike Held, Jakob M. Burgstaller, Johann Steurer.

Writing – review & editing: Ulrike Held, Jakob M. Burgstaller, Maria M. Wertli, Giuseppe Pichierri, Sebastian Winklhofer, Florian Brunner, François Porchet, Mazda Farshad, Johann Steurer.

References

1. Irwin ZN, Hilibrand A, Gustavel M, McLain R, Shaffer W, Myers M, et al. Variation in surgical decision making for degenerative spinal disorders. Part I: lumbar spine. Spine (Phila Pa 1976). 2005 Oct 01; 30 (19):2208–13. PMID: 16205348.
2. Weinstein JN, Lurie JD, Olson PR, Bronner KK, Fisher ES. United States' trends and regional variations in lumbar spine surgery: 1992–2003. Spine (Phila Pa 1976). 2006 Nov 01; 31(23):2707–14. https://doi.org/10.1097/01.brs.0000248132.15231.fe PMID: 17077740. Pubmed Central PMCID: PMC2913862.
3. Delitto A, Piva SR, Moore CG, Fritz JM, Wisniewski SR, Josbeno DA, et al. Surgery versus nonsurgical treatment of lumbar spinal stenosis: a randomized trial. Ann Intern Med. 2015 Apr 07; 162(7):465–73. https://doi.org/10.7326/M14-1420 PMID: 25844995.
4. Malmivaara A, Slatis P, Heliovaara M, Sainio P, Kinnunen H, Kankare J, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. Spine (Phila Pa 1976). 2007 Jan 01; 32(1):1–8. https://doi.org/10.1097/01.brs.0000251014.81875.6d PMID: 17202885.
5. Weinstein JN, Tosteson TD, Lurie JD, Tosteson A, Blood E, Herkowitz H, et al. Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. Spine (Phila Pa 1976). 2010 Jun 15; 35(14):1329–38. https://doi.org/10.1097/BRS.0b013e3181e0f04d PMID: 20453723. Pubmed Central PMCID: PMC3392200.
6. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Blood E, Hanscom B, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. N Engl J Med. 2008 Feb 21; 358(8):794–810. https://doi.org/10.1056/NEJMoa0707136 PMID: 18287602. Pubmed Central PMCID: PMC2576513.
7. Slatis P, Malmivaara A, Heliovaara M, Sainio P, Herno A, Kankare J, et al. Long-term results of surgery for lumbar spinal stenosis: a randomised controlled trial. Eur Spine J. 2011 Jul; 20(7):1174–81. https://doi.org/10.1007/s00586-010-1652-y PMID: 21240530. Pubmed Central PMCID: PMC3175822.
8. Lurie JD, Tosteson TD, Tosteson A, Abdu WA, Zhao W, Morgan TS, et al. Long-term outcomes of lumbar spinal stenosis: eight-year results of the Spine Patient Outcomes Research Trial (SPORT). Spine (Phila Pa 1976). 2015 Jan 15; 40(2):63–76. https://doi.org/10.1097/BRS.0000000000000731 PMID: 25569524. Pubmed Central PMCID: PMC4288009.
9. Atlas SJ, Keller RB, Wu YA, Deyo RA, Singer DE. Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the maine lumbar spine study. Spine (Phila Pa 1976). 2005 Apr 15; 30(8):936–43. PMID: 15834339.

10. Steurer J, Nydegger A, Held U, Brunner F, Hodler J, Porchet F, et al. LumbSten: The lumbar spinal stenosis outcome study. Bmc Musculoskel Dis. 2010 Nov 2; 11. PubMed PMID: WOS:000284325600001. English.
11. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj-Brit Med J. 2015 Jan 7; 350. PubMed PMID: WOS:000348130300002. English.
12. von Elm E, Altman DG, Pocock SJ, Gotzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Brit Med J. 2007 Oct 20; 335(7624):806–8. PubMed PMID: WOS:000250630700036. English. https://doi.org/10.1136/bmj.39335.541782.AD PMID: 17947786
13. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol. 1995 Nov; 48 (11):1369–78. PMID: 7490600. Epub 1995/11/01. eng.
14. Tuli SK, Yerby SA, Katz JN. Methodological approaches to developing criteria for improvement in lumbar spinal stenosis surgery. Spine (Phila Pa 1976). 2006 May 15; 31(11):1276–80. https://doi.org/10.1097/01.brs.0000217615.20018.6c PMID: 16688044. Epub 2006/05/12. eng.
15. Zucherman JF, Hsu KY, Hartjen CA, Mehalic TF, Implicito DA, Martin MJ, et al. A multicenter, prospective, randomized trial evaluating the X STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication: two-year follow-up results. Spine (Phila Pa 1976). 2005 Jun 15; 30(12):1351–8. PMID: 15959362. Epub 2005/06/17. eng.
16. Hansraj KK, O'Leary PF, Cammisa FP Jr., Hall JC, Fras CI, Cohen MS, et al. Decompression, fusion, and instrumentation surgery for complex lumbar spinal stenosis. Clin Orthop Relat Res. 2001 Mar (384):18–25. PMID: 11249164. Epub 2001/03/16. eng.
17. Fokter SK, Yerby SA. Patient-based outcomes for the operative treatment of degenerative lumbar spinal stenosis. Eur Spine J. 2006 Nov; 15(11):1661–9. https://doi.org/10.1007/s00586-005-0033-4 PMID: 16369827. Epub 2005/12/22. eng.
18. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. Spine (Phila Pa 1976). 1996 Apr 1; 21(7):796–803. PMID: 8779009.
19. Miettinen OS. Up from 'Clinical Epidemiology'& EBM. Up from Clinical Epidemiology & EBM: Springer; 2010. p. 15–20.
20. Athiviraham A, Wali ZA, Yen D. Predictive factors influencing clinical outcome with operative management of lumbar spinal stenosis. Spine J. 2011 Jul; 11(7):613–7. https://doi.org/10.1016/j.spinee.2011.03.008 PMID: 21482198.
21. Aalto T, Sinikallio S, Kroger H, Viinamaki H, Herno A, Leinonen V, et al. Preoperative predictors for good postoperative satisfaction and functional outcome in lumbar spinal stenosis surgery—a prospective observational study with a two-year follow-up. Scand J Surg. 2012; 101(4):255–60. https://doi.org/10.1177/145749691210100406 PMID: 23238500.
22. Aalto TJ, Malmivaara A, Kovacs F, Herno A, Alen M, Salmi L, et al. Preoperative predictors for postoperative clinical outcome in lumbar spinal stenosis: systematic review. Spine (Phila Pa 1976). 2006 Aug 15; 31(18):E648–63. https://doi.org/10.1097/01.brs.0000231727.88477.da PMID: 16915081.
23. Sigmundsson FG, Kang XP, Jonsson B, Stromqvist B. Prognostic factors in lumbar spinal stenosis surgery A prospective study of imaging- and patient-related factors in 109 patients who were operated on by decompression. Acta Orthop. 2012 Oct; 83(5):536–42. PubMed PMID: WOS:000310015700017. English. https://doi.org/10.3109/17453674.2012.733915 PMID: 23083437
24. WHO. Body mass index—BMI [October 2017]. Available from: http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi.
25. Ng LCL, Tafazal S, Sell P. The effect of duration of symptoms on standard outcome measures in the surgical treatment of spinal stenosis. European Spine Journal. 2007 Feb; 16(2):199–206. PubMed PMID: WOS:000244192200005. English. https://doi.org/10.1007/s00586-006-0078-z PMID: 16496190
26. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale—An updated literature review. J Psychosom Res. 2002 Feb; 52(2):69–77. PubMed PMID: WOS:000173932200003. English. PMID: 11832252
27. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014 Aug 1; 35(29):1925-+. PubMed PMID: WOS:000342232100007. English. https://doi.org/10.1093/eurheartj/ehu207 PMID: 24898551
28. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating: Springer Science & Business Media; 2008.

29. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007; 16(3):219–42. PubMed PMID: WOS:000247467200003. English. https://doi.org/10.1177/0962280206074463 PMID: 17621469

30. Dunkler D, Sauerbrei W, Heinze G. Global, Parameterwise and Joint Shrinkage Factor Estimation. J Stat Softw. 2016 Mar; 69(8):1–19. PubMed PMID: WOS:000373918200001. English.

31. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. Biometrika. 1999 Dec; 86(4):948–55. PubMed PMID: WOS:000084833000018. English.

32. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. p. http://www.R-project.org/.

33. Localio A, Goodman SN, Meibohm A, et al. STatistical code to support the scientific story. Annals of Internal Medicine. 2018.

34. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007 Mar 15; 165(6):710–8. PubMed PMID: WOS:000244655200014. English. https://doi.org/10.1093/aje/kwk052 PMID: 17182981

35. Kreiner DS, Shaffer WO, Baisden JL, Gilbert TJ, Summers JT, Toton JF, et al. An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). Spine Journal. 2013 Jul; 13(7):734–43. PubMed PMID: WOS:000321491100006. English. https://doi.org/10.1016/j.spinee.2012.11.059 PMID: 23830297

36. Pearson A, Lurie J, Tosteson T, Zhao WY, Abdu W, Weinstein JN. Who Should Have Surgery for Spinal Stenosis? Treatment Effect Predictors in SPORT. Spine. 2012 Oct 1; 37(21):1791–802. PubMed PMID: WOS:000309550600012. English. https://doi.org/10.1097/BRS.0b013e3182634b04 PMID: 23018805

37. Hey HWD, Luo N, Chin SY, Lau ETC, Wang P, Kumar N, et al. The Predictive Value of Preoperative Health-Related Quality-of-Life Scores on Postoperative Patient-Reported Outcome Scores in Lumbar Spine Surgery. Global Spine Journal. 2017:2192568217701713.

38. Adamova B, Vohanka S, Dusek L, Jarkovsky J, Chaloupka R, Bednarik J. Outcomes and their predictors in lumbar spinal stenosis: a 12-year follow-up. European Spine Journal. 2015 Feb; 24(2):369–80. PubMed PMID: WOS:000349437200023. English. https://doi.org/10.1007/s00586-014-3411-y PMID: 24943641

39. Gulati S, Nordseth T, Nerland US, Gulati M, Weber C, Giannadakis C, et al. Does daily tobacco smoking affect outcomes after microdecompression for degenerative central lumbar spinal stenosis?—A multicenter observational registry-based study. Acta Neurochir. 2015 Jul; 157(7):1157–64. PubMed PMID: WOS:000356341900011. English. https://doi.org/10.1007/s00701-015-2437-1 PMID: 25943982

40. Paulsen RT, Bouknaitir JB, Fruensgaard S, Carreon L, Andersen M. Prognostic Factors for Satisfaction After Decompression Surgery for Lumbar Spinal Stenosis. Neurosurgery. 2017.

41. Sinikallio S, Aalto T, Lehto SM, Airaksinen O, Herno A, Kroger H, et al. Depressive symptoms predict postoperative disability among patients with lumbar spinal stenosis: A two-year prospective study comparing two age groups. Disabil Rehabil. 2010; 32(6):462–8. PubMed PMID: WOS:000275534200005. English. https://doi.org/10.3109/09638280903171477 PMID: 19849649

42. Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating: Springer Science & Business Media; 2010.

43. Mannion AF, Denzler R, Dvorak J, Grob D. Five-year outcome of surgical decompression of the lumbar spine without fusion. European Spine Journal. 2010 Nov; 19(11):1883–91. PubMed PMID: WOS:000284684000011. English. https://doi.org/10.1007/s00586-010-1535-2 PMID: 20680372